



TRABAJO DE GRADO
Opción Seminario-Diplomado.

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN EL RESULTADO Y PROYECCIÓN DE ADOPCIÓN DE MASCOTAS DEL
REFUGIO AMOR Y HUELLAS, UTILIZANDO ESTRATEGIAS DE MACHINE
LEARNING**

Corporación Universitaria Remington.
Nombre de la facultad: Especialización
Nombre del programa académico: Especialización en Dirección de Operaciones y Mejoramiento
Continuo

Estudiante:
Tatiana Ortiz Marin
Tutor: Juan Carlos Briñez de León
Opción de Trabajo de grado Seminario-Diplomado.
2024.

Tabla de Contenidos

Contenido

Resumen.....	3
Palabras clave.....	3
Marco conceptual y contextual	4
Pregunta problema	7
Acercamiento a los datos	7
Descripción de variables	8
Aproximaciones con gráficos – analítica.....	10
Objetivos.....	19
Desarrollo e implementación del aprendizaje.....	19
Preparación de los datos.....	25
Modelo de toma de decisiones	27
Validación del modelo	30

Resumen

En la mayoría de las ciudades y estados de todos los países, los perros y gatos callejeros se vuelven una parte importante del paisaje mostrando sus enfermedades y sufrimiento constante. Es bien sabido, que muchas de estas mascotas han tenido diferentes dueños quienes los abandonan a su suerte por diferentes motivos y muchos otros han nacido en la calle, producto de los padres abandonados.

Esta investigación busca encontrar resultados en las preferencias de mascotas al momento de adoptar un nuevo integrante en las familias estadounidenses, siendo estos animales domésticos entre perros, gatos, conejos y pericos que son las opciones disponibles en el refugio estudiado. Al entender estas preferencias, se puede realizar un pronóstico de probabilidad de adopción en cada grupo de mascotas.

Es importante con estos estudios poder sensibilizar a las familias sobre la importancia de adoptar animales y brindar un hogar definitivo que además de garantizar el bienestar del animal, se garantiza también el bienestar físico y emocional en la familia adoptante, ya que, a través de muchos estudios se ha demostrado como la influencia y presencia de una mascota puede controlar episodios delicados de salud, así como estrés, ansiedad y depresión que son los trastornos principales del siglo XXI, evitando también con la compañía de mascotas el desarrollo de enfermedades terminales y difíciles de tratar.

Palabras clave

Mascotas, adopción, refugios de mascotas, salud mental, bienestar animal, responsabilidad, Machine Learning, Conexión, compañía.

Marco conceptual y contextual

Es muy común en cualquier lugar del mundo, encontrar en las calles animales en estado de abandono que sufren maltrato y rechazo diariamente, además de aumentar su población por reproducciones entre ellos. Según algunos estudios, muchos de estos animales, haciendo un énfasis en perros y gatos que corresponden al volumen más alto, han tenido dueños en sus inicios quienes optaron por utilizarlos como medios de ingresos económicos a través de la venta de cachorros, explotando así al animal y llevándolo a condiciones extremas de insalubridad debido al constante maltrato.

Muchos otros animales han sido abandonados por familias que se mudan de casa o cambian de País y no dudan en dejar las mascotas a su suerte. Otros tantos, abandonan los animales por crecimiento en sus familias, es decir nacimiento de hijos quienes pueden llegar con condiciones de salud o alergias que les impida estar cerca de ciertos animales.

Es bien sabido que diariamente en las calles se encuentran animales muertos o que causan accidentes ya que están desorientados o en su defecto buscando comida para sobrevivir.

Actualmente, la importancia que se le da a los animales ha incrementado en un 100% respecto a lo que se veía en años anteriores, donde no eran tenidos en cuenta como un miembro de la familia. Se han realizado campañas desde los gobiernos y los grupos animalistas donde recalcan la tenencia responsable de mascotas, así como también se

empezó a castigar la violencia contra ellos, siendo desde mi punto de vista un gran avance en la humanidad.

Entre algunas de las ayudas que han brindado los gobiernos en diferentes países para el cuidado de mascotas está: Mejorar el bienestar animal por medio de prestaciones de servicios veterinarios que contribuyan a mejorar la calidad de vida de los animales y a su vez de la ciudadanía.

Otro beneficio está en brindar jornadas gratuitas de esterilización que abarca animales abandonados, así como mascotas de familias de escasos recursos que no pueden pagar este tipo de cirugías, evitando la reproducción irresponsable de animales y sobrepoblación de perros y gatos que queden en condición de calle.

Realizar capacitaciones en las ciudades acerca de la tenencia responsable de mascotas involucrando conocimientos acerca de las condiciones de salud de los animales, su desarrollo en las diferentes etapas de la vida y las necesidades que tienen dentro de un entorno familiar. (Jaramillo, 2016)

La sobrepoblación de animales de compañía es un problema creado por los humanos, con importantes costos para los mismos, y que solo puede ser abordado a través de acciones humanas. Debido al inmenso volumen de animales sin hogar, los refugios destinados a su acogida frecuentemente se encuentran sobrepasados en sus recursos. Así, típicamente, estos animales viven en ambientes estériles y ruidosos, con pocas oportunidades de desplegar comportamientos propios de la especie (Fratkin, 2013)

Otro estudio enfocado en el tema de adopción indaga por las percepciones de un grupo de personas que van a adoptar una nueva mascota -gatos o perros- y también las percepciones o cambios en el estado emocional y físico entendiendo si se vio influenciado por la presencia de las mascotas en sus vidas. Se encuentra con esta investigación que las personas que se referían a las mascotas como un animal, los buscaban como objeto de apoyo y responsabilidad de un miembro de la familia mientras que aquellos que los veían como un miembro más, buscaban estas adopciones para encontrar compañía y amor y así mismo, este último grupo mostró después de tres meses mayor impacto en el bienestar emocional. (Taborda, 2018)

Es interesante ver en las investigaciones que algunos países tienen segmentados los perros según su condición, denominándolos abandonados o asilvestrados.

Los primeros están en el grupo de animales abandonados que no son peligrosos y suelen alimentarse de desechos en las calles. El segundo grupo está compuesto por perros que dejaron de ser domésticos y generalmente se ven en jauría en zonas rurales. Estos perros desarrollan la habilidad de caza y suelen alimentarse de ganados u otros animales, afectando a los propietarios de estos. También en ocasiones resultan peligrosos para la comunidad ya que si sienten alguna mínima amenaza pueden atacar a las personas o niños causando graves lesiones o infecciones.

Existen evidencias de que ciertos comportamientos anómalos como la coprofagia, el acicalamiento excesivo y los movimientos estereotipados o repetitivos, son típicamente expresados bajo condiciones de cautiverio. Las condiciones de vida de los refugios caninos

pueden considerarse análogas a la de animales salvajes viviendo en cautiverio. En este sentido, una de las alteraciones conductuales observadas en perros de refugio es la alta frecuencia de acicalamiento, especialmente en aquellos que se encuentran aislados socialmente. (Gabriela Barrera, 2008)

Uno de los problemas que afrontan los países actualmente es la cantidad de animales abandonados en las calles, padeciendo enfermedades, causando accidentes en las vías y excediendo el cupo en los refugios destinados a la protección de derechos de estos.

Los refugios están constantemente incentivando la adopción a través de campañas con los gobiernos y alcaldías disminuyendo la ocupación del refugio y mejorando la calidad de vida de los animales.

Los algoritmos de machine learning pueden apoyar para la proyección de datos sobre la posibilidad de adopción de las mascotas en determinados refugios, entendiendo las preferencias de las familias en cuanto tamaño y color del animal.

Pregunta problema

¿Cómo desarrollar una estrategia computacional para ayudar en la predicción de adopción de mascotas y tenencia responsable, a partir de algoritmos de Machine Learning?

Acercamiento a los datos

Los datos para establecer la probabilidad de adopción de mascotas en el refugio Amor y Huellas fueron tomados de la plataforma Kaggle publicados por Rabie El Kharoua con el nombre “Predict Pet Adoption Status Dataset”, el cual se enfoca en estadísticas sobre la adopción de mascotas de un refugio en particular que tiene como alternativas pericos, conejos, perros y gatos. Con base en los resultados puede generar estrategias e impulsar la adopción de animales segmentando familias, entendiendo sus intereses y presupuesto en la tarifa de adopción entregando la mejor sugerencia para la familia y la mascota. Permite identificar el porcentaje de posibilidad que uno de los animales sea adoptado. Esto puede verse influenciado por la edad, color, tamaño y antecedentes médicos.



Imagen 1. Conjunto de datos.

Descripción de variables

El dataset seleccionado para este trabajo cuenta con 12 columnas compuestas por once entradas y una salida que corresponde a la decisión y/o probabilidad de adopción de las mascotas. Las variables contenidas en la base de datos se pueden evidenciar en la siguiente imagen.

```

#Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2007 entries, 0 to 2006
Data columns (total 13 columns):
#   Column              Non-Null Count  Dtype
---  -
0   PetID                2007 non-null  int64
1   PetType              2007 non-null  object
2   Breed                2007 non-null  object
3   AgeMonths            2007 non-null  int64
4   Color                2007 non-null  object
5   Size                 2007 non-null  object
6   WeightKg             2007 non-null  object
7   Vaccinated           2007 non-null  int64
8   HealthCondition      2007 non-null  int64
9   TimeInShelterDays   2007 non-null  int64
10  AdoptionFee          2007 non-null  int64
11  PreviousOwner        2007 non-null  int64
12  AdoptionLikelihood   2007 non-null  int64
dtypes: int64(8), object(5)
memory usage: 204.0+ KB

```

Imagen 2. Lista de variables.

Como se puede evidenciar en la imagen anterior, esta base de datos cuenta con 2007 registros donde 5 variables contienen información en texto y los 7 restantes son números enteros.

Asimismo, se realiza un estimado de los valores mínimos o máximos en las variables para entender hasta dónde llega la información teniendo un primer análisis de datos. Esta información explica lo siguiente: se encuentra que el mayor número de días de las mascotas en el refugio es de 89 días y lo mínimo es 1 día. Se encuentra también que para la tarifa de adopción el mayor valor es 499 usd y lo mínimo es un pago de 0 usd. Otro dato importante es la edad de las mascotas, donde el mayor cuenta con 179 meses de edad y el menor cuenta con 1 mes de edad. Se anexa imagen con detalle.

	PetID	AgeMonths	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood
count	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000
mean	1503.000000	92.279522	0.701046	0.196313	43.974091	249.142003	0.301943	0.328351
std	579.515315	52.148363	0.457914	0.397307	25.740253	142.887040	0.459215	0.469730
min	500.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
25%	1001.500000	48.000000	0.000000	0.000000	21.000000	127.000000	0.000000	0.000000
50%	1503.000000	94.000000	1.000000	0.000000	45.000000	242.000000	0.000000	0.000000
75%	2004.500000	138.000000	1.000000	0.000000	66.000000	375.000000	1.000000	1.000000
max	2506.000000	179.000000	1.000000	1.000000	89.000000	499.000000	1.000000	1.000000

Imagen 3. Análisis de los datos.

Estos datos también pueden generar una idea de la edad que tienen las mascotas en el refugio y como esta variable podría influir en la decisión de adopción, ya que un perro adulto, es decir mayor de 3 años, no sería la primera opción para las familias teniendo en cuenta que estos podrían tener costumbres no aceptadas en determinados hogares. Asimismo, cuando superan los 7 años ya tendrían un riesgo mayor de desarrollar enfermedades que pueden tener un costo alto de atención en las veterinarias.

(KHAROUA, 2024)

Aproximaciones con gráficos – analítica

Corriendo los datos en a través de machine learning, se obtienen gráficas que nos muestran resultados y porcentajes según los conceptos que se quieran analizar y que sean relevantes para la decisión de adopción. En la primera gráfica que se estudia, se entiende el porcentaje que representa cada animal en el refugio Amor y Huellas. Para este caso, el

animal con mayor representación es el perro teniendo un 26.0%, aunque se tiene un resultado muy parejo entre las cuatro opciones de mascota. Ver imagen 4.

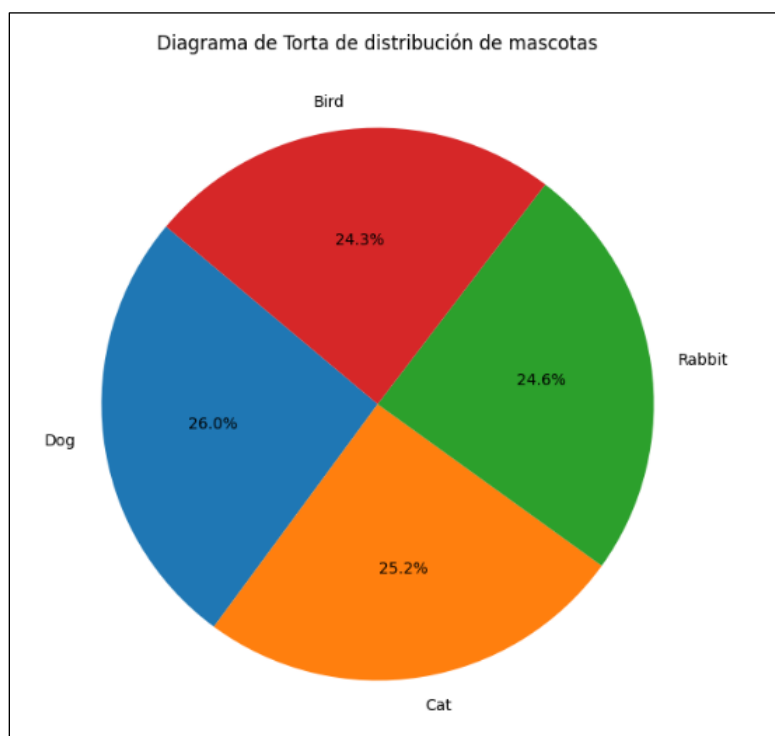


Imagen 4. Distribución de tipo de mascota en el refugio

Asimismo, se realiza una gráfica que indica aproximadamente la cantidad de animales que hay por cada categoría indicando incluso la raza de este. En el caso de los pericos y los conejos solo se segmentan en un tipo ya que no tienen una raza asociada, contando con el mayor número de registros, es decir entre 490 y 500 animales en cada una de las dos categorías. Para el caso de los gatos, se tienen dos razas que son Siamese y Persian los cuales están muy parejos en unidad de animales oscilando en 260. Para el último caso, se cuenta con tres razas de perros los cuales tienen una diferencia más

notoria, siendo el labrador la raza con mayor representación con 200 perros. Ver imagen

5.

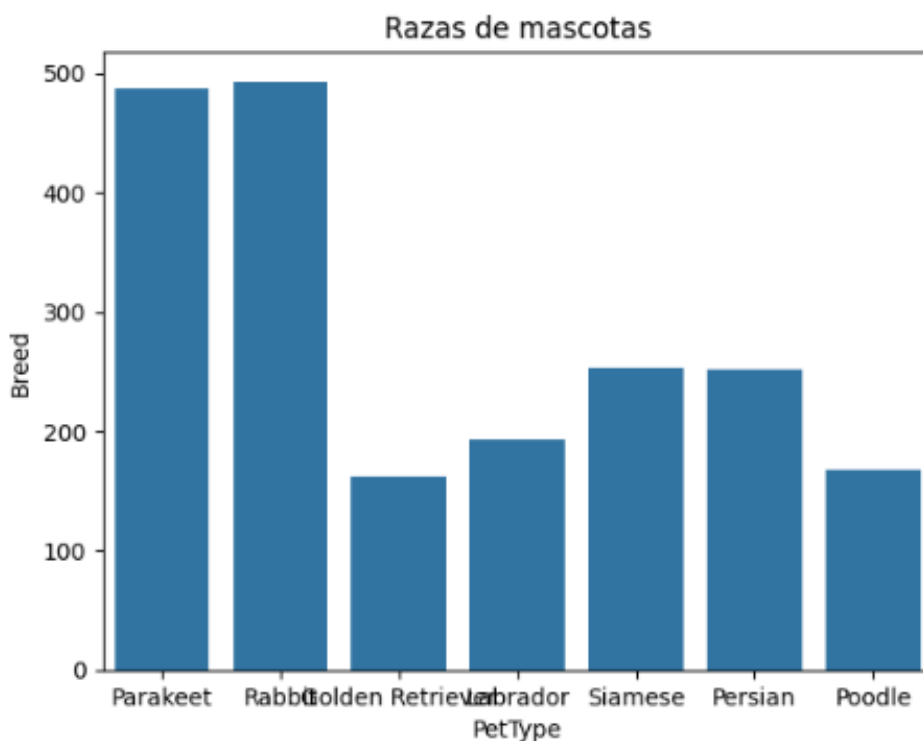


Imagen 5. Distribución de raza de mascotas

Otro aspecto crucial en la decisión de adopción para los refugios es el color del animal, puesto que esta característica en algunas ocasiones puede verse asociada a paradigmas y creencias locativas, especialmente con mascotas de color negro las cuales están con un riesgo más alto según algunas investigaciones de ser maltratados, abandonados o incluso sacrificados por grupos que practican actividades de ocultismo.

(Videla, 2020)

En la gráfica que se muestra a continuación, se puede apreciar que el color también está equitativamente distribuido entre las mascotas oscilando entre 19% y 20% teniendo el mayor porcentaje el color de pelaje blanco con un 20,9%.

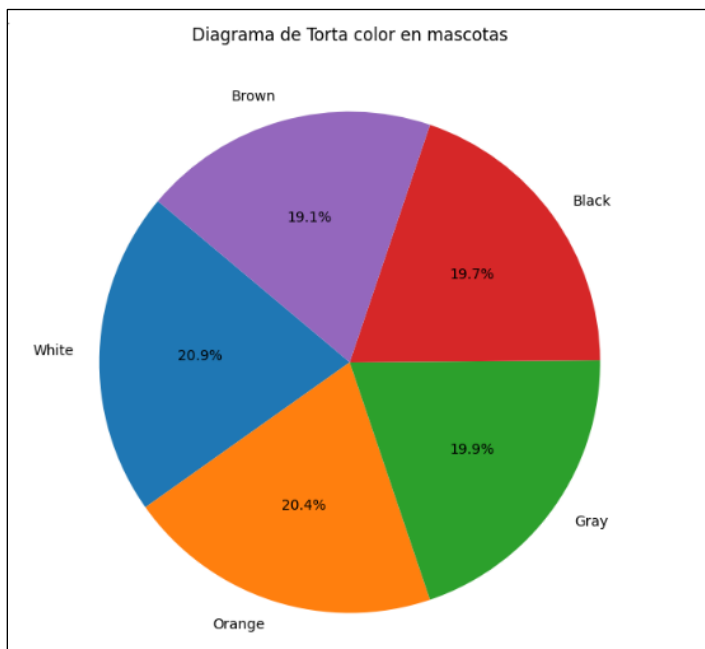
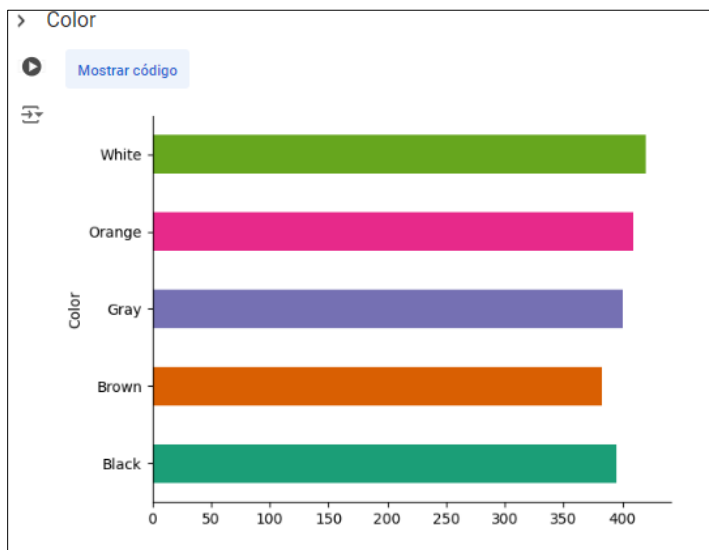


Imagen 6. Porcentaje de color



A su vez, se genera un diagrama de torta para entender el porcentaje de condición de la salud en las mascotas. Según el dataset, si la condición tiene valor 1, es porque sufre alguna patología o desarrolla enfermedades que pueden afectar la probabilidad de adopción. Aunque la diferencia es alta, el porcentaje para animales con enfermedades tiene un valor importante que indica que se tienen muchas mascotas en el refugio que requieren posiblemente tratamiento médico o medicinas para mejorar su calidad de vida y preservarla por más tiempo. El porcentaje de animales con enfermedad es de un 19.6% lo que sugiere que se tiene una cantidad de 393.3 en dicha condición médica, siendo un número considerable para el presupuesto que tiene un refugio por lo cual no pueden ser bien atendidas y genera el riesgo alto de transmitir estas enfermedades a los demás animales.

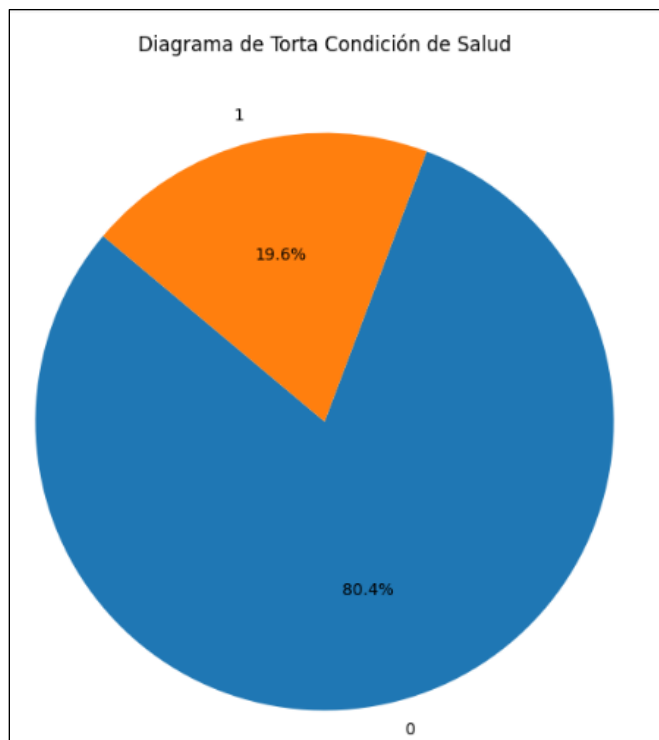


Imagen 7. Condición de salud

Otro análisis importante para tener en cuenta a la hora de graficar es el tiempo de permanencia en los refugios de estos animales, donde a través de un diagrama de caja se puede observar que la mayoría de las mascotas llevan entre 20 y 60 días allí esperando una oportunidad de tener una familia. Bajo este mismo modelo de gráfica, se conoce el promedio que predomina para las tarifas de adopción. Ver imagen 8 y 9.

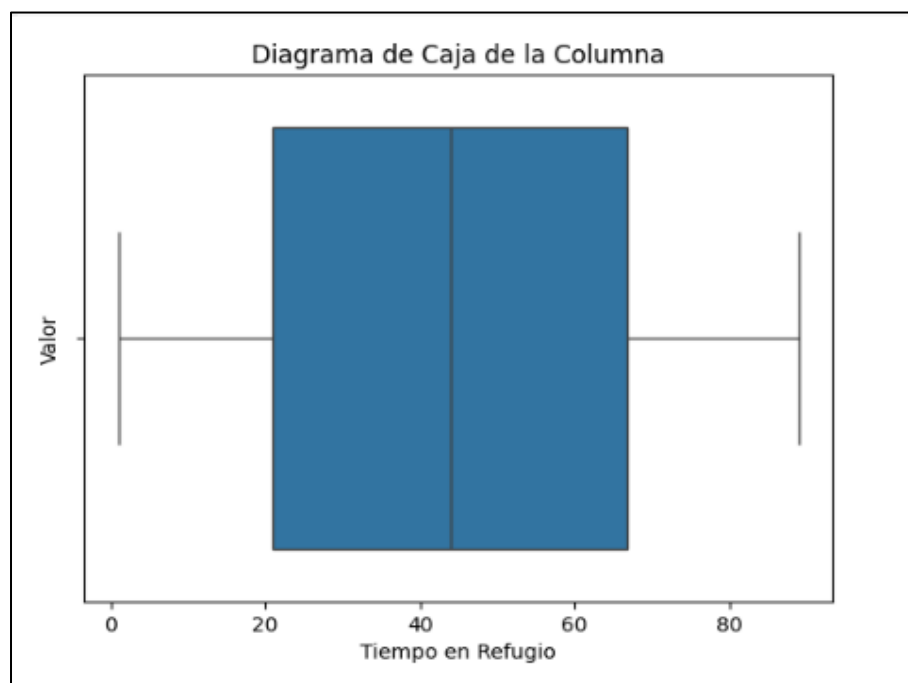


Imagen 8. Diagrama de días en refugio

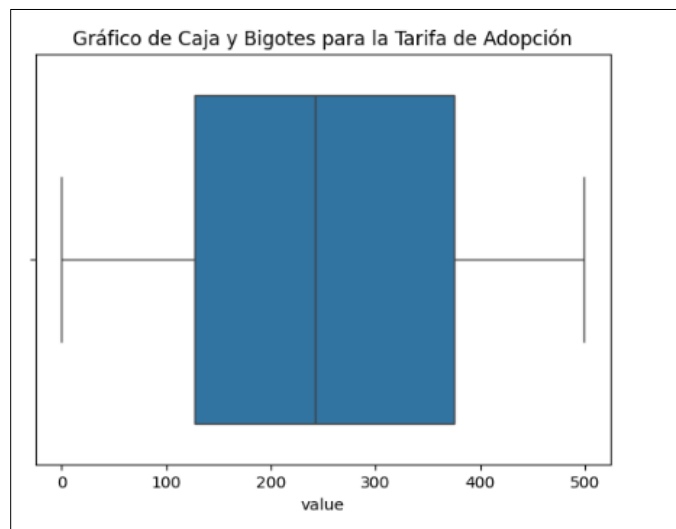


Imagen 9.

Se busca identificar la intensidad entre las variables “tarifa de adopción” y “tiempo en refugio” entendiendo si tienen alguna relación y según el resultado obtenido se entiende que no hay ninguna correlación, ya que la tarifa de adopción impuesta para la mascota no depende del tiempo que lleve en el refugio.

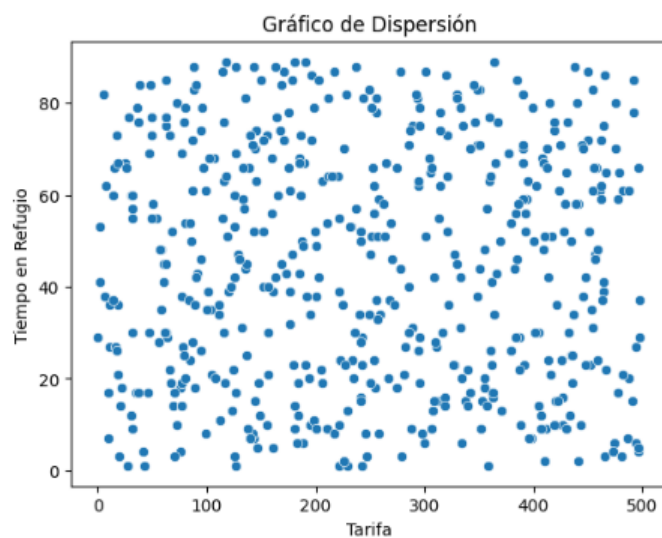


Imagen 9. Relación

Se realizan más análisis a través de gráficas con ayuda de machine learning, donde se quiere identificar el rango que predomina para cada variable. En este caso, para entender la distribución de la tarifa de adopción, los días en el refugio y la edad de las mascotas se usaron gráficas de frecuencia y se encuentra lo siguiente: para la tarifa, el mayor porcentaje se ubica en el rango de 99,8 usd a 199,6 usd con aproximadamente 495 animales. Para el análisis de la edad, el rango predominante es entre 143.4 y 161.2 meses de edad contando con aproximadamente 280 animales en este rango. Por último, el tiempo de permanencia en el refugio para el cual, la mayoría de las mascotas se ubican entre 1 y 18,6 días. Para el detalle, observar imágenes 10, 11 y 12.

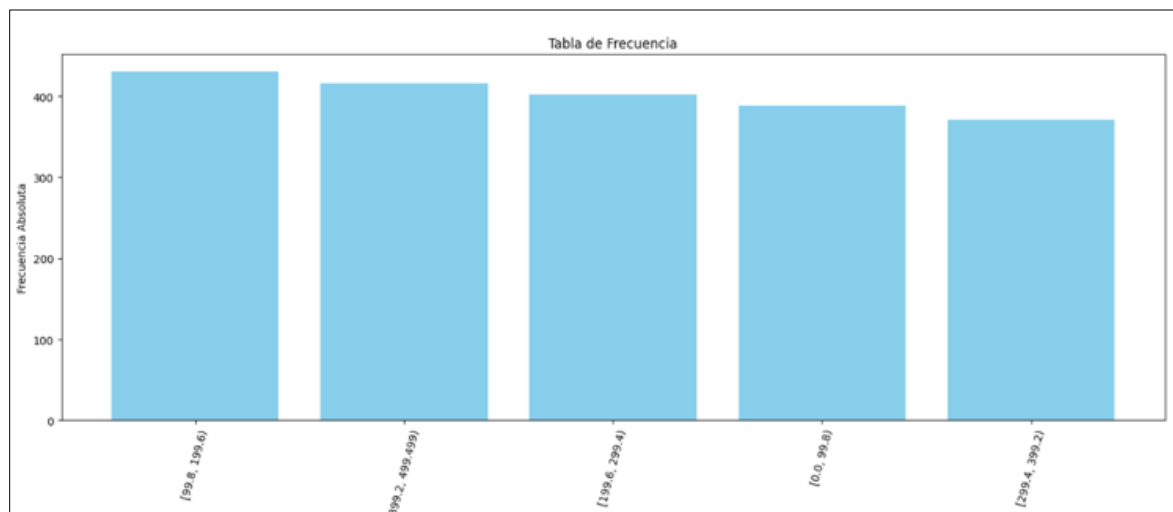


Imagen 10. Tabla de frecuencia tarifa de adopción

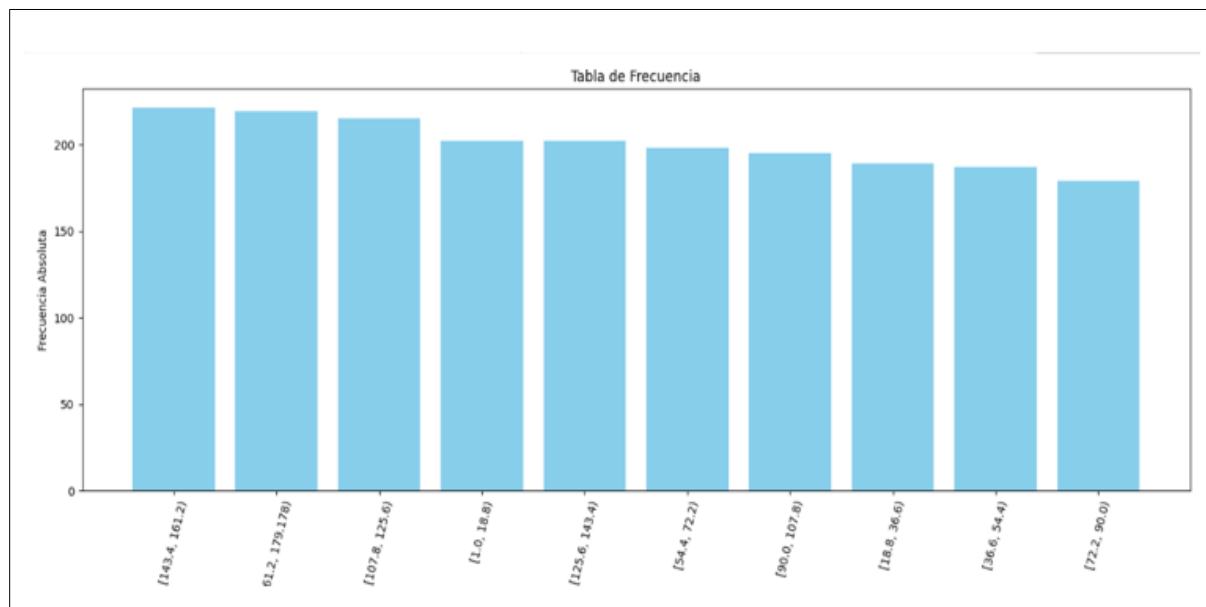


Imagen 11. Tabla de frecuencia edad de la mascota

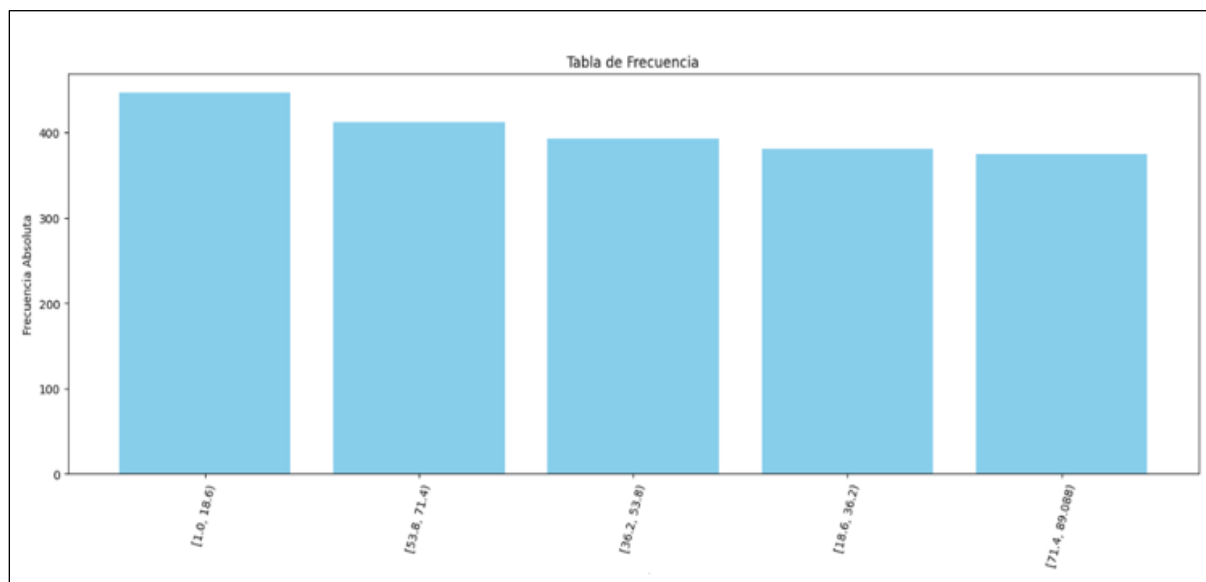


Imagen 12. Tabla de frecuencia tiempo en refugio

Objetivos

Objetivo general.

Implementar un algoritmo computacional para el análisis y toma de decisiones en la proyección de adopción de mascotas del refugio Amor y Huellas, utilizando estrategias de machine learning.

Objetivos específicos.

- Caracterizar y procesar los datos de interés, con miras a la toma de decisiones informadas.
- Implementar un algoritmo de Machine learning para la toma de decisiones a partir de los datos de interés.
- Evaluar y analizar el desempeño de los algoritmos implementados para la toma de decisiones.
- Validar el funcionamiento de toma de decisiones a partir de datos nuevos.

Desarrollo e implementación del aprendizaje

Para el desarrollo de este trabajo, inicialmente se selecciona la base de datos de interés desde la plataforma Kaggle. Se buscan informes, investigaciones y libros desde Google Scholar evitando plagios o información poco confiable. Una vez escogido el tema, se descarga en archivo CSV para trabajarlo a través de librerías de Python. Se

desarrollan los pasos a continuación para lograr información limpia y exacta que permita analizar las variables y generar una asertiva toma de decisiones.

En primer lugar, se carga el archivo para conocer los datos. Imagen 13.

```
[25] #Para cargar los datos
import pandas as pd
from google.colab import files
uploaded = files.upload()
for filename in uploaded.keys():
    Conjunto_Datos = pd.read_csv(filename, sep=',')
Conjunto_Datos.head(10)
```

Elegir archivos: pet_adoption_data.csv

- pet_adoption_data.csv(text/csv) - 145369 bytes, last modified: 5/7/2024 - 100% done

Saving pet_adoption_data.csv to pet_adoption_data.csv

	PetID	PetType	Breed	AgeMonths	Color	Size	WeightKg	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood
0	500	Bird	Parakeet	131	Orange	Large	5.039.767.822.529.510	1	0	27	140	0	0
1	501	Rabbit	Rabbit	73	White	Large	16.086.726.854.616.700	0	0	8	235	0	0
2	502	Dog	Golden Retriever	136	Orange	Medium	20.762.862.789.067.600	0	0	85	385	0	0
3	503	Bird	Parakeet	97	White	Small	3.339.423.254.344.140	0	0	61	217	1	0
4	504	Rabbit	Rabbit	123	Gray	Large	2.049.809.976.164.300	0	0	28	14	1	0
5	505	Dog	Labrador	70	Brown	Large	20.986.260.902.397.900	0	0	87	301	1	0
6	506	Bird	Parakeet	169	Brown	Small	1.090.261.279.164.480	1	0	70	440	1	0
7	507	Cat	Siamese	13	Orange	Large	7.252.683.192.048.290	1	0	3	137	0	1
8	508	Bird	Parakeet	49	Brown	Medium	2.459.759.785.196.190	1	1	69	405	0	0
9	509	Bird	Parakeet	60	Gray	Large	72.959.936.546.268.200	0	0	73	231	1	0

Imagen 13. Cargue de datos

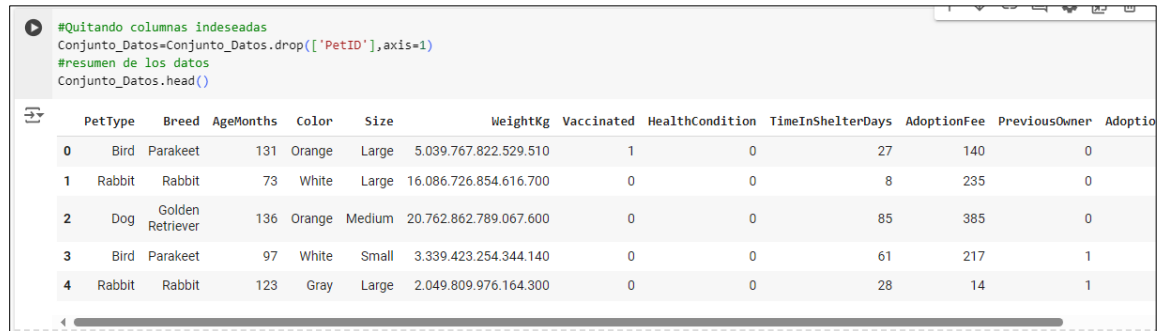
Posteriormente se corre el código que me permite visualizar un análisis rápido de los datos y los valores mínimos, máximos o promedio de las variables numéricas.

```
#Análisis de los datos
Conjunto_Datos.describe()
```

	PetID	AgeMonths	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood
count	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000	2007.000000
mean	1503.000000	92.279522	0.701046	0.196313	43.974091	249.142003	0.301943	0.328351
std	579.515315	52.148363	0.457914	0.397307	25.740253	142.887040	0.459215	0.469730
min	500.000000	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
25%	1001.500000	48.000000	0.000000	0.000000	21.000000	127.000000	0.000000	0.000000
50%	1503.000000	94.000000	1.000000	0.000000	45.000000	242.000000	0.000000	0.000000
75%	2004.500000	138.000000	1.000000	0.000000	66.000000	375.000000	1.000000	1.000000
max	2506.000000	179.000000	1.000000	1.000000	89.000000	499.000000	1.000000	1.000000

Imagen 14. Análisis de los datos

Después de conocer los promedios, variables y tipo de dato, se procede a eliminar las columnas que no agregan valor o son irrelevantes para la toma de decisiones. Para este caso en particular se elimina inicialmente la columna de PetID.



```
#Quitando columnas indeseadas
Conjunto_Datos=Conjunto_Datos.drop(['PetID'],axis=1)
#resumen de los datos
Conjunto_Datos.head()
```

	PetType	Breed	AgeMonths	Color	Size	WeightKg	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	Adoptio
0	Bird	Parakeet	131	Orange	Large	5.039.767.822.529.510	1	0	27	140	0	
1	Rabbit	Rabbit	73	White	Large	16.086.726.854.616.700	0	0	8	235	0	
2	Dog	Golden Retriever	136	Orange	Medium	20.762.862.789.067.600	0	0	85	385	0	
3	Bird	Parakeet	97	White	Small	3.339.423.254.344.140	0	0	61	217	1	
4	Rabbit	Rabbit	123	Gray	Large	2.049.809.976.164.300	0	0	28	14	1	

Imagen 15. Quitando columnas

Se cambian algunas columnas que contienen texto para volverlas numéricas y tener la facilidad de graficar las variables y obtener resultados. Datos importantes para considerar en la toma de decisiones serían la variable de tamaño “Size” y la variable “Color” ya que son características que influyen directamente en la probabilidad de adopción de la mascota. En la variable Size, se modifica de la siguiente forma: al texto small se le asigna el número 0, para médium se asigna el número 1 y para large se asigna el número 2. Así se muestra en la imagen 16.

Del mismo modo, se realiza modificación para obtener gráficas a la variable de color de la siguiente forma: Orange se asigna el número 0, para White se asigna el número 1, para gray se asigna el número 2, para Brown se asigna el número 3 y para black se asigna el número 4. Ver imagen 17.

```
# Cambiando rango de valores del dataframe
Estrategia_1={'Small':0,'Medium':1,'Large':2}
Conjunto_Datos['Size']=Conjunto_Datos['Size'].map(Estrategia_1)
Conjunto_Datos.head()
```

	PetType	Breed	AgeMonths	Color	Size	WeightKg	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	Adopt
0	Bird	Parakeet	131	Orange	2	5.039.767.822.529.510	1	0	27	140	0	0
1	Rabbit	Rabbit	73	White	2	16.086.726.854.616.700	0	0	8	235	0	0
2	Dog	Golden Retriever	136	Orange	1	20.762.862.789.067.600	0	0	85	385	0	0
3	Bird	Parakeet	97	White	0	3.339.423.254.344.140	0	0	61	217	1	1
4	Rabbit	Rabbit	123	Gray	2	2.049.809.976.164.300	0	0	28	14	1	1

Imagen 16. Cambio de rango de valor Size

Próximos pasos: [Generar código con Dataset](#) [Ver gráficos recomendados](#)

```
# Corregir errores de formato (por ejemplo, convertir a minúsculas)
Dataset['Color']=Dataset['Color'].str.lower()
Dataset.head()
#Analizando opciones de variable categórica
Dataset['Color'].unique()

array(['orange', 'white', 'gray', 'brown', 'black'], dtype=object)

#Convirtiendo a numérico
Conversion = {'orange':0,'white':1,'gray':2,'brown':3,'black':4}
Dataset['Color']=Dataset['Color'].replace(Conversion)
Dataset.head(7)
```

	PetType	Breed	AgeMonths	Color	Size	Weightkg	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood
0	Bird	Parakeet	131	0	Large	5.039.767.822.529.510	1	0	27	140	0	0
1	Rabbit	Rabbit	73	1	Large	16.086.726.854.616.700	0	0	8	235	0	0
2	Dog	Golden Retriever	136	0	Medium	20.762.862.789.067.600	0	0	85	385	0	0
3	Bird	Parakeet	97	1	Small	3.339.423.254.344.140	0	0	61	217	1	0
4	Rabbit	Rabbit	123	2	Large	2.049.809.976.164.300	0	0	28	14	1	0
5	Dog	Labrador	70	3	Large	20.986.260.902.397.900	0	0	87	301	1	0
6	Bird	Parakeet	169	3	Small	1.090.261.279.164.480	1	0	70	440	1	0

Imagen 17. Cambio de rango de valor Color

Otro cambio importante que se debe considerar para graficar y tomar decisiones es el tipo de animal, ya que esta es la primera característica que se tiene en cuenta para seleccionar un adoptante. Se modifican los rangos de la siguiente forma: a la variable Bird se le asigna el número 0, para Rabbit se asigna el número 1, para Dog se asigna el número 2 y para Cat se asigna el número 3.

```

#Convirtiendo a numérico
Conversion = {'Bird':0, 'Rabbit':1, 'Dog':2, 'cat':3}
Dataset['PetType']=Dataset['PetType'].replace(Conversion)
Dataset.head(7)

```

	PetType	AgeMonths	Size	Weightkg	Vaccinated	Healthcondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood
0	0	131	0	5.039.767.822.529.510	1	0	27	140	0	0
1	1	73	0	16.086.726.854.616.700	0	0	8	235	0	0
2	2	136	1	20.762.862.789.067.600	0	0	85	385	0	0
3	0	97	2	3.339.423.254.344.140	0	0	61	217	1	0
4	1	123	0	2.049.809.976.164.300	0	0	28	14	1	0
5	2	70	0	20.986.260.902.397.900	0	0	87	301	1	0
6	0	169	2	1.090.261.279.164.480	1	0	70	440	1	0

Imagen 18. Cambiando rango de valor del tipo de mascota

Con estos cambios se pueden obtener gráficos adicionales que permitan entender las relaciones entre las variables y la dependencia de una con la otra. Luego de modificar los rangos de texto a número se genera una gráfica de correlación que indica una relación más alta entre la probabilidad de adopción y la vacunación en la mascota con un resultado de 0.3.

Se encuentra muy poca relación entre la información de mascotas que han tenido dueño anteriormente contra los días de permanencia en el refugio con un resultado de 0.008, al igual que la relación entre los días en el refugio y la probabilidad de adopción. Para esto, se muestra la imagen 19.

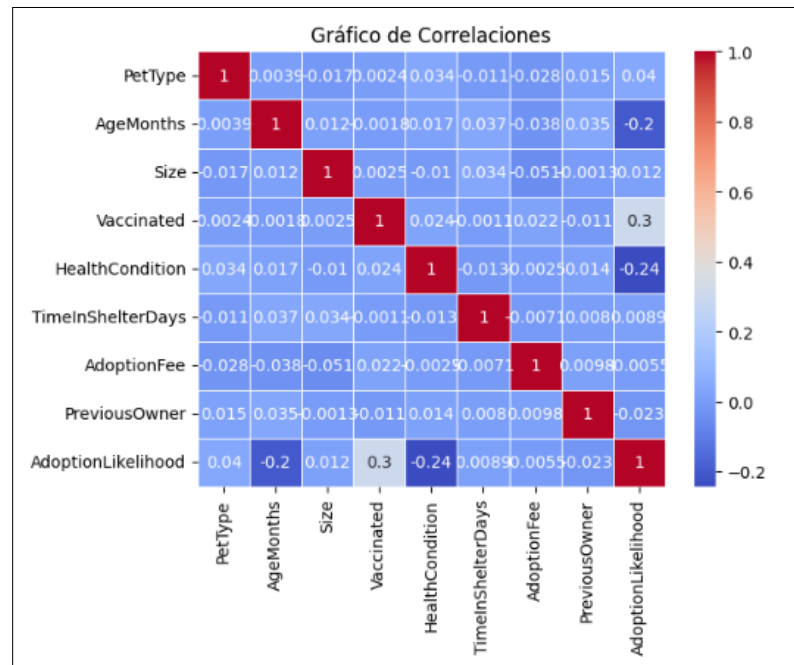


Imagen 19



Asimismo, se genera un gráfico de clustering de resultados con gráfico de dispersión y se segmentan muy notoriamente. No hay relación entre ellos. Ver imagen 21.

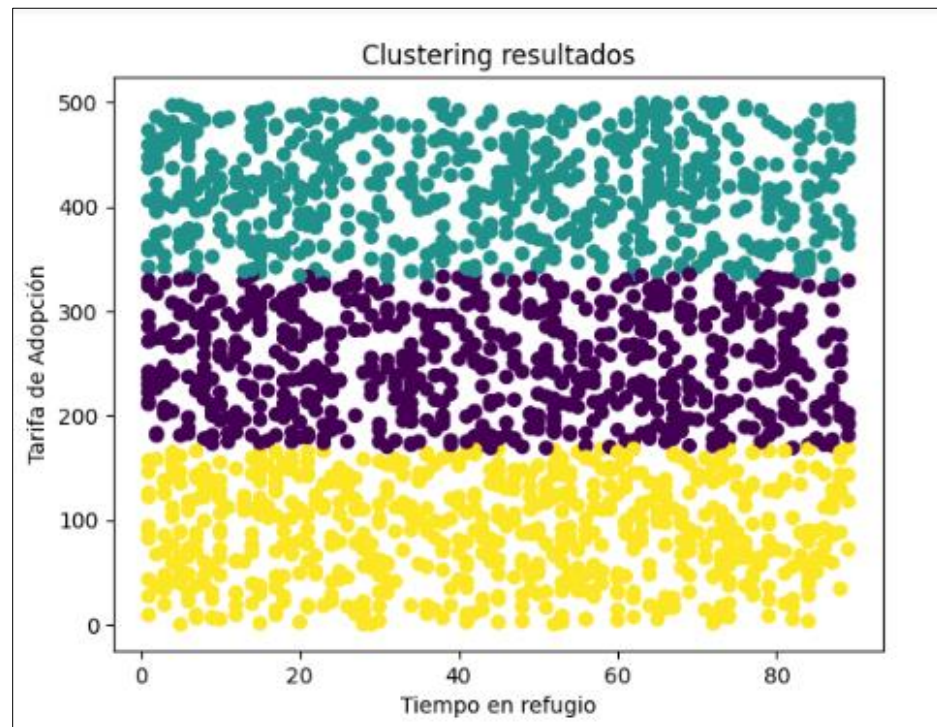


Imagen 21.

Preparación de los datos

Como se demostró en el capítulo anterior, se inició con eliminar columnas que no aportan valor y no son relevantes para la toma de decisiones. En el caso de este dataset solo aplica la columna PetID. Posteriormente se modificaron los rangos de valor texto para ponerlos numéricos y poder obtener información confiable y exacta de los porcentajes de las variables.

En este paso se vuelven numéricas las variables de tamaño, color y tipo de mascota, asignando números de 0 a 4 para identificar.

También se hizo una limpieza de errores quitando filas con valores faltantes, como se muestra en la siguiente imagen:

```
[ ] # Identificar valores faltantes
print(Dataset.isnull().sum())

# Eliminar filas con valores faltantes
Dataset = Dataset.dropna()
Dataset.head(7)
```

⇒	PetID	0
	PetType	0
	AgeMonths	0
	Size	0
	WeightKg	0
	Vaccinated	0
	HealthCondition	0
	TimeInShelterDays	0
	AdoptionFee	0
	PreviousOwner	0
	AdoptionLikelihood	0
	dtype:	int64

Imagen 23. Quitando errores

Asimismo, se eliminan duplicados como se evidencia en el capítulo anterior que no se tenía ninguno. Quedan los mismos 2007 registros iniciales del archivo.

```
▶ # Identificar y eliminar duplicados
print('Número de filas antes de eliminar duplicados:', len(Dataset))
Dataset = Dataset.drop_duplicates()
print('Número de filas después de eliminar duplicados:', len(Dataset))
```

⇒	Número de filas antes de eliminar duplicados:	2007
	Número de filas después de eliminar duplicados:	2007

Imagen 22. Eliminación de duplicados

Después de limpiar los datos, se procede a generar las gráficas que se quieren analizar y los informes respecto a las variables y correlación entre ellas. Las gráficas de resultados son expuestas en el capítulo anterior con comentarios y análisis al respecto. El archivo queda habilitado después de correr la librería de códigos para realizar el modelo de toma de decisiones.

Modelo de toma de decisiones

Para el trabajo y el tipo de información, se desarrolla el modelo de clustering a través de centroides donde se pretende crear grupos o clientes nuevos segmentando sus necesidades y desde el refugio determinar según sus ingresos y gustos qué mascota podría adquirir. Este desarrollo es sumamente importante, ya que, con el clustering, además de entregar la mejor opción, se está llevando a cabo un tema de responsabilidad social con los animales puesto que se garantiza una vida digna y segura a partir del momento de la adopción.

El sistema genera los grupos 0, 1 y 2 los cuales segmentan necesidades y presupuesto de clientes (este último para la tarifa de adopción de la mascota), también tiene en cuenta el tipo de cuidado que puede dar la familia adoptante y el espacio con el que cuenta para tener determinado tipo de animal. Según los análisis del refugio se ofrecen las siguientes alternativas: Si la persona según su encuesta queda en el grupo 0, se le ofrecen únicamente adoptar un conejo. Si la persona queda asignada en el grupo 1 aplica para adoptar un perro y si queda en el grupo 2 se le ofrece adoptar un gato.

Tal como se muestra en el desarrollo de Python, al correr los códigos, se hizo la prueba con tres clientes diferentes aplicando cada uno para uno de los tres animales disponibles. Para la opción de los pericos no va a ser tan viable ofrecer adopción ya que son animales de fauna silvestre quienes podrían entrar en un proceso de rehabilitación desde el refugio y posteriormente con apoyo de las entidades gubernamentales, llevarlos a su hábitat natural mejorando la calidad de vida.

El paso siguiente es la predicción de grupos en el conjunto de datos para generar centroides que los segmenta en determinados rangos por grupo. El sistema asignó los grupos 0, 1 y 2. En la imagen 20 se muestra el desarrollo del código y los rangos que asignó a cada grupo creado.

```
# Importando las bibliotecas necesarias
import numpy as np
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Transformando el dataset a formato numérico
Dataset_numerico = np.array(Dataset)

# Elegir el número de grupos (Dependiendo de las necesidades del usuario)
k = 3

#Generando y entrenando el modelo
Modelo_cluster = KMeans(n_clusters=k, n_init = 'auto', random_state=42)
Modelo_cluster.fit(Dataset_numerico)

#Prediciendo etiquetas de grupos en el conjunto de datos
Predicciones = Modelo_cluster.fit_predict(Dataset_numerico)

# Asignando etiqueta como columna nueva al dataset
Dataset_2 = Dataset.copy()
Dataset_2['Grupo'] = Predicciones
Dataset_2.head()
```

Imagen 20. Creación de grupos

```

#Prediciendo etiquetas de grupos en el conjunto de datos
Predicciones = Modelo_Cluster.fit_predict(Dataset_numerico)

# Asignando etiqueta como columna nueva al dataset
Dataset_2 = Dataset.copy()
Dataset_2['Grupo'] = Predicciones
Dataset_2.head()

```

	PetType	AgeMonths	Size	Vaccinated	HealthCondition	TimeInShelterDays	AdoptionFee	PreviousOwner	AdoptionLikelihood	Grupo
0	0	131	0	1	0	27	140	0	0	2
1	1	73	0	0	0	8	235	0	0	0
2	2	136	1	0	0	85	385	0	0	1
3	0	97	2	0	0	61	217	1	0	0
4	1	123	0	0	0	28	14	1	0	2

Imagen 20. Creación de grupos

```

Centroides:
  PetType  AgeMonths    Size  Vaccinated  HealthCondition \
0  1.503693  94.630724  1.042836   0.686854      0.184638
1  1.497682  88.795981  0.952087   0.720247      0.202473
2  1.559297  93.248902  1.051245   0.696925      0.202050

  TimeInShelterDays  AdoptionFee  PreviousOwner  AdoptionLikelihood
0      43.995569    249.636632      0.304284      0.344165
1      43.670788    418.916538      0.310665      0.319938
2      44.240117     87.825769      0.291362      0.320644

```

Imagen 20. Creación de grupos

Validación del modelo

```

▶ # Ingresar datos del nuevo almacén (9 variables)
Clientes_New=np.zeros((1,9))
Clientes_New[0,0]=0 #Tipo de mascota
Clientes_New[0,1]=131 #Edad mascota
Clientes_New[0,2]=2 #Tamaño
Clientes_New[0,3]=0 #vacunación
Clientes_New[0,4]=1 #Condición de salud
Clientes_New[0,5]=80 #Tiempo en refugio
Clientes_New[0,6]=1 #Propietario antiguo
Clientes_New[0,7]=0 #Probabilidad adopción
Clientes_New[0,8]=268 #Tarifa

# Realizar la predicción del cluster del nuevo almacen
Etiqueta_nueva = modelo_cargado.predict(Clientes_New)
print('')
print("El nuevo usuario pertenece al cluster:", Etiqueta_nueva)
if Etiqueta_nueva == 0:
    print('Aplica a adopción de conejo')

if Etiqueta_nueva == 1:
    print('Aplica a adopción de perro')

if Etiqueta_nueva == 2:
    print('Aplica a adopción de gato')

```



 El nuevo usuario pertenece al cluster: [2]
 Aplica a adopción de gato

Imagen 23. Validación de modelo

En esta primera tabla de resultado donde se corre la librería con el código, se evidencia que el nuevo cliente aplica para una adopción de gato con las variables importantes de tiempo en refugio de 80 días, edad de la mascota de 131 meses y con una tarifa de adopción de 268 dólares.

```

▶ # Ingresar datos del nuevo almacén (9 variables)
Clientes_New=np.zeros((1,9))
Clientes_New[0,0]=2 #Tipo de mascota
Clientes_New[0,1]=95 #Edad mascota
Clientes_New[0,2]=1 #Tamaño
Clientes_New[0,3]=1 #vacunación
Clientes_New[0,4]=0 #Condición de salud
Clientes_New[0,5]=44 #Tiempo en refugio
Clientes_New[0,6]=249 #Tarifa
Clientes_New[0,7]=0 #Propietario antiguo
Clientes_New[0,8]=0 #Probabilidad de adopción

# Realizar la predicción del cluster del nuevo almacen
Etiqueta_nueva = modelo_cargado.predict(Clientes_New)
print('')
print("El nuevo usuario pertenece al cluster:", Etiqueta_nueva)
if Etiqueta_nueva == 0:
    print('Aplica a adopción de conejo')

if Etiqueta_nueva == 1:
    print('Aplica a adopción de perro')

if Etiqueta_nueva == 2:
    print('Aplica a adopción de gato')

```

↔

```

El nuevo usuario pertenece al cluster: [0]
Aplica a adopción de conejo

```

Imagen 24. Validación de modelo

Tal como en el primer ensayo del modelo, para este cliente con los datos suministrados aplica la adopción de un conejo, quedando en el grupo número 0. Los datos suministrados por este cliente es que la mascota lleve en el refugio 44 días, que tenga una edad no superior a 95 meses. Para un conejo, esta sería una edad ya muy adulta, entendiendo que según estudios el promedio de vida de un conejo con condiciones de vida saludables puede estar entre 7 a 9 años. Adicional, ofrece presupuesto de adopción de 249 dólares.

```

▶ # Ingresar datos del nuevo almacén (9 variables)
Clientes_New=np.zeros((1,9))
Clientes_New[0,0]=3 #Tipo de mascota
Clientes_New[0,1]=169 #Edad mascota
Clientes_New[0,2]=0 #Tamaño
Clientes_New[0,3]=0 #vacunación
Clientes_New[0,4]=1 #Condición de salud
Clientes_New[0,5]=70 #Tiempo en refugio
Clientes_New[0,6]=440 #Tarifa
Clientes_New[0,7]=0 #Propietario antiguo
Clientes_New[0,8]=0 #Probabilidad de adopción

# Realizar la predicción del cluster del nuevo almacen
Etiqueta_nueva = modelo_cargado.predict(Clientes_New)
print('')
print("El nuevo usuario pertenece al cluster:", Etiqueta_nueva)
if Etiqueta_nueva == 0:
    print('Aplica a adopción de conejo')

if Etiqueta_nueva == 1:
    print('Aplica a adopción de perro')

if Etiqueta_nueva == 2:
    print('Aplica a adopción de gato')

```


 El nuevo usuario pertenece al cluster: [1]
 Aplica a adopción de perro

Imagen 25. Validación de modelo

En la tercera validación de modelo, con un cliente ofreciendo condiciones diferentes, se incluye en el grupo 1, el cual indica una adopción de perro. Para este caso, las condiciones suministradas por el cliente fueron: edad mascota 169 meses, tarifa de adopción de 440 dólares y con un tiempo en el refugio no superior a 70 días.

Conclusiones y trabajos futuros

- En el trabajo desarrollado se materializó la construcción de un algoritmo para la toma de decisiones informadas en la proyección de la probabilidad de adopción de mascotas en un refugio estadounidense a través de la utilización de algoritmos de machine learning.
- En cuanto a los modelos utilizados, el algoritmo de mejor desempeño fue el de clustering y segmentación de clientes, ya que brinda la oportunidad de ofrecer mascotas según las condiciones de vida de la familia garantizando el bienestar para ambas partes.
- Para trabajos futuros se proyecta la implementación del modelo directamente en la industria animal, específicamente en albergues con mascotas abandonadas, entendiendo la relación entre el abandono y el desarrollo de enfermedades, así como cantidad de población con mascotas en los países versus cantidad de ese grupo con mascotas adoptadas estimando el porcentaje de adopción de cada estado y mejorando las alternativas y condiciones de los refugios y animales en estado de abandono.
- Al hacer una revisión general de la situación en que se encuentran cientos de animales abandonados es imprescindible pensar en una legislación que los proteja y que por medio de la educación hacia las personas especialmente a las nuevas

generaciones, esta problemática cambie, aumentando la sensibilización en las personas y dejando claro los temas científicos e investigaciones que demuestran como un animal doméstico puede influir positivamente en el desarrollo de los niños y personas con alguna patología, además de disminuir los índices de trastorno de depresión y ansiedad.

- Este trabajo tiene como fin promover la tenencia responsable no solo para quienes gustan de los animales sino para todos, ya que es algo que afecta a la población nacional y una legislación a favor de esto podrá ayudar a promover estos cambios.

Referencias

- Fratkin, J. L. (2013). *The role of coat color and ear shape on the perception of personality in dogs*. Obtenido de Fratkin, J. L., & Baker, S. C. (2013). The role of coat color and ear shape on the perception of personality in dogs. *Anthrozoös*, 26(1), 125-133. doi: 10.2752/175303713X13534238631632
- Gabriela Barrera, A. J. (2008). *CALIDAD DE VIDA EN PERROS ALOJADOS EN REFUGIOS*. Universidad de Buenos Aires.
- Jaramillo, M. R. (2016). *Adopción y tenencia responsable de mascotas*. Santiago de Chile: Universidad de Chile, Escuela de Periodismo.
- KHAROUA, R. E. (2024). *Predict Pet Adoption Status Dataset*. Obtenido de Kaggle: <https://www.kaggle.com/datasets/rabieelkharoua/predict-pet-adoption-status-dataset>
- Taborda, M. L. (2018). *Revista Universidad Eafit*, pág. 3.
- Videla, M. D. (29 de 02 de 2020). Síndrome del perro negro. *Revista Argentina de Ciencias del Comportamiento*, pág. 6.